ADVANCES IN NEURODEGENERATIVE AND PSYCHIATRIC IMAGING SPECIAL FEATURE: FULL PAPER

Optimisation and usefulness of quantitative analysis of $^{18}$F-florbetapir PET

1,2Daniel Fakhry-Darian, 1,2Neva Hiten Patel, 2Sairah Khan, 2Tara Barwick, 2William Svensson, 2Sameer Khan, 2Richard J Perry, 3,4Paresh Malhotra, 3,4Christopher J Carswell, 1,Kuldip S Nijran and 2Zarni Win

1Radiological Sciences Unit, Imperial College Healthcare NHS Trust, Charing Cross Hospital, Fulham Palace Road, London, UK
2Department of Nuclear Medicine, Imperial College Healthcare NHS Trust, Charing Cross Hospital, Fulham Palace Road, London, UK
3Department of Neurology, Imperial College Healthcare NHS Trust, Charing Cross Hospital, Fulham Palace Road, London, UK
4Division of Brain Sciences, Imperial College London, UK

Address correspondence to: Mr Daniel Fakhry-Darian
E-mail: daniel.fakhry-darian@uhb.nhs.uk

**Objectives:** This study investigates the usefulness of quantitative SUVR thresholds on sub types of typical (type A) and atypical (non-type A) positive (Aβ+) and negative (Aβ-) $^{18}$F-florbetapir scans and aims to optimise the thresholds.

**Methods:** Clinical $^{18}$F-florbetapir scans ($n = 100$) were categorised by sub type and visual reads were performed independently by three trained readers. Inter-reader agreement and reader-to-reference agreement were measured. Optimal SUVR thresholds were derived by ROC analysis and were compared with thresholds derived from a healthy control group and values from published literature.

**Results:** Sub type division of $^{18}$F-florbetapir PET scans improves accuracy and agreement of visual reads for type A: accuracy 90%, 96% and 70% and agreement $\kappa > 0.7$, $\kappa \geq 0.85$ and $-0.1 < \kappa < 0.9$ for all data, type A and non-type A respectively. Sub type division also improves quantitative classification accuracy of type A: optimum mcSUVR thresholds were found to be 1.32, 1.18 and 1.48 with accuracy 86%, 92% and 76% for all data, type A and non-type A respectively.

**Conclusions:** Aβ+/Aβ- mcSUVR threshold of 1.18 is suitable for classification of type A studies (sensitivity = 97%, specificity = 88%). Region-wise SUVR thresholds may improve classification accuracy in non-type A studies. Amyloid PET scans should be divided by sub type before quantification.

**Advances in knowledge:** We have derived and validated mcSUVR thresholds for Aβ+/Aβ- $^{18}$F-florbetapir studies. This work demonstrates that division into sub types improves reader accuracy and agreement and quantification accuracy in scans with typical presentation and highlights the atypical presentations not suited to global SUVR quantification.

INTRODUCTION

Amyloid brain PET/CT imaging is a powerful tool in the investigation of suspected Alzheimer’s disease (AD) – $^{18}$F-florbetapir binds to beta-amyloid (Aβ) plaques in the brain, the presence of which can support a diagnosis of AD. These studies may contribute significantly to making an earlier diagnosis, play an important role in the investigation of patients with cognitive impairment and dementia, especially in those with atypical and mixed clinical presentations, and have been shown to change patient diagnosis and management.1

Several studies agree that quantifying the uptake of florbetapir in Aβ specific brain regions, relative to a region of non-specific binding, can aid visual reporting of $^{18}$F-florbetapir PET/CT investigations. Uptake thresholds in these specific binding regions can be used to aid identification of amyloid positive (Aβ+) and negative (Aβ-) patients.2-7 Previous studies which have investigated these Aβ+/Aβ- thresholds have not considered varying patterns of uptake. In this study we investigate quantitative thresholds according to varying scan sub types defined by the pattern of tracer uptake.

Scan sub-types

The method for reporting $^{18}$F-florbetapir scans is visual delineation of areas with high and low tracer uptake, especially in the differentiation of the white matter-grey matter
interface (WMGM). Loss of contrast at the WMGM indicates a moderate or high level of amyloid plaque deposition which is compatible with a neuropathological diagnosis of AD. However, this loss is not always global and may be localised or display other patterns of uptake. To aid reporting we have further classified $^{18}$F-florbetapir PET/CT studies into type A (with typical features) or non-type A (sub types B-D, with atypical features) for both positive and negative scans (Figure 1).

Figure 1. Imaging characteristics of the sub groups of scans defined in this study with examples for comparison: a) sub type A-D positive and b) sub type A-D negative $^{18}$F-florbetapir examinations.
Positive type A scans have complete loss of WMGM differentiation in all cortical regions globally. Positive scans which may have just two cortical regions with loss of WMGM differentiation are termed type B, those with one or more areas in which grey matter activity is intense and clearly exceeds activity in adjacent white matter are positive type C (Figure 1a).

Negative type A scans have clear WMGM differentiation in all cortical regions, where the grey matter has no or low tracer uptake and is identical to that of cerebellar grey matter. Type B negative scans are uncommon and have good WMGM differentiation in all cortical regions, but tracer uptake in the cerebral grey matter is elevated relative to the cerebellum. Type C negative scans demonstrate one cortical region with loss of WMGM differentiation (Figure 1b). Type D scans have been identified as those with poor signal to noise ratio (SNR) or marked atrophy, however the sample used in this study does not have any scans of this classification.

Quantification of $^{18}$F-florbetapir
A publication by Avid Radiopharmaceuticals suggests that a mean cortical standardised uptake value ratio (mcSUVR) threshold of 1.17 correlated with the Consortium to Establish a Registry for Alzheimer’s Disease (CERAD) plaque count for definite $\text{A}^\beta$ and “intermediate or high likelihood” for the NIA/Reagan Institute method. Application of this mcSUVR threshold as an adjunct to visual reporting demonstrated an improvement in inter-reader variability, however the authors highlight the mcSUVR was based on a small sample ($n = 19$) and is also not suitable to apply to individual cortical regions.

$\text{A}^\beta+/\text{A}^\beta-$ threshold of 1.10 achieved 97% sensitivity, 100% specificity and 98% accuracy for detection of “moderate to frequent neuritic plaques” (CERAD) in 59 patients. In a later study the same group used ROC curve analysis for threshold optimisation and again calculated mcSUVR threshold of 1.10 and replicated their previous sensitivity and specificity results on a larger sample. This threshold was cross-calibrated with Hermes BRASS, a clinically approved software package, yielding mcSUVR = 1.11. The methodology for SUVR measurements used by Pontecorvo et al is the same as the methodology used here (the same software package).

Studies for the Alzheimer’s Disease Neuroimaging Initiative (ADNI) have compared quantitative PET measurements and CSF measurements of amyloid plaques (A$\beta$). They found that mcSUVR and CSF-A$\beta$ changes were closely correlated and that using a threshold mcSUVR of 1.11 was consistent with CSF-A$\beta$ changes were closely correlated and that

The mcSUVR values in the literature demonstrate correlation with autopsy data and CSF-A$\beta$ changes and improving inter-reader variability, however none of these studies have classified scans into sub types as have been defined at our centre. Similarly, cortical ROI SUVR have been published for other software and methodologies, but so far no regional cortical thresholds have been published for the method and software used here. Considering that a single cortical region showing an area of intense uptake is an indication of a positive amyloid scan, albeit not that common an appearance, it would be useful to investigate meaningful thresholds for regional quantification which may be beneficial to support diagnosis of non-type A scans.

AIM
The aims of this study are twofold:

1. To assess the usefulness of $\text{A}^\beta+/\text{A}^\beta-$ thresholds derived from a healthy control population when classifying type A and non-type A scans relative to a reference standard.
2. To optimise the mcSUVR and regional SUVR for different scan sub types.

METHODS
At the time of the study over 200 clinical $^{18}$F-florbetapir studies had been acquired at our centre. The first 100 were selected for this study (age range: 41–88 years, median: 68 years). All scans were independently reviewed by two trained readers who identified sub type and classified the scans as $\text{A}^\beta+$ or $\text{A}^\beta-$. The consensus of the two reports was chosen as the reference standard, similar to Schreiber et al.

Scan classifications and sub types are given in Table 1. Table B and C were grouped into “non-type A” scans as these represent atypical uptake patterns and both groups had relatively small samples. However as the definitions of type B and C are different they have been analysed separately and included in the tables.

Image acquisition
Patients were injected with 370 MBq $^{18}$F-florbetapir (333 MBq min.) with imaging at 40 min post injection. Each study was acquired as a 20 minute list mode scan on a Siemens Biograph 64 PET/CT system at Charing Cross-Hospital, London. Frames with excessive motion were left out of the final reconstruction and a minimum of 10 minutes of data was reconstructed as a static image with OSEM 2D (4 iterations 14 subsets), 3 mm gaussian filter and 168 × 168 matrix.

<table>
<thead>
<tr>
<th>Type</th>
<th>$\text{A}^\beta+$</th>
<th>$\text{A}^\beta-$</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>46</td>
<td>54</td>
<td>100</td>
</tr>
<tr>
<td>Type A</td>
<td>38</td>
<td>41</td>
<td>79</td>
</tr>
<tr>
<td>Non-type A</td>
<td>8</td>
<td>13</td>
<td>21</td>
</tr>
<tr>
<td>Type B</td>
<td>8</td>
<td>7</td>
<td>15</td>
</tr>
<tr>
<td>Type C</td>
<td>0</td>
<td>6</td>
<td>6</td>
</tr>
</tbody>
</table>
Trained reader review
All 100 studies were independently reviewed by three experienced reviewers (X, Y and Z), all of whom received reader training from the radiopharmaceutical manufacturer (Avid Radiopharmaceuticals, Eli Lilly) and had 3 + years’ experience in reporting amyloid PET/CT.

18F-Florbetapir Quantification
Quantitative evaluation was performed using Hermes BRASS (Hermes Medical Imaging Solutions). The chosen study was registered to a PET template averaged from 80 healthy control studies. ROIs were applied in six cortical regions of the brain: frontal medial orbital cortex, anterior cingulate, lateral temporal lobes, precuneus, posterior cingulate and parietal lobe (the cortical regions used were provided by Avid and are the same used by Pontecorvo). The florbetapir uptake in each cortical region was normalised to the uptake in the whole cerebellum (standardised uptake value ratio, SUVR). Using the cerebellum to normalise the uptake correlates well with changes in amyloid pathology. The mean of the six SUVR values was then calculated (mean cortical SUVR, mcSUVR).

A detailed guide to visual and quantitative interpretation of 18F-florbetapir images can be found in the summary of product characteristics (SPC). According to the SPC, a positive scan will have either:

1. Two or more brain areas in which there is reduced or absent GMWM contrast, or
2. One or more areas in which grey matter activity is intense and clearly exceeds activity in adjacent white matter.

This study investigated two metrics to identify a positive scan in BRASS, (i) mcSUVR: this metric is commonly reported in published literature so allows for comparison with other studies. The \( A\beta^+/A\beta^- \) threshold was initially set to two standard deviations (SD) above the mcSUVR calculated from the template of healthy controls. This threshold value is referred to as the global (i.e. using all 6 ROIs) healthy control (gHC) threshold; (ii) Two-region classification: if two of the six cortical ROIs each have SUVR ≥ the healthy control SUVR +2 SD the scan was classed as \( A\beta^+ \). The healthy control SUVR +2 SD is referred to as the regional (i.e. just for 1 ROI) healthy control (rHC) threshold.

Statistics and analysis
The \( \kappa \) score, \( \kappa(\pm95\% \text{ CI}) \), was used to assess agreement using classifications as given by Altman.19 Confusion matrices were calculated for each reader and for the quantitative evaluation relative to the reference standard and divided into sub types. ROC curves were used to define potential SUVR thresholds and the area under the ROC curve (AUC) was used to compare classification of scan sub types. Confusion matrices for optimised mcSUVR and ROI SUVR thresholds were compared to the gHC and rHC thresholds using the Matthews Correlation Coefficient (MCC).20

RESULTS
Inter-reader agreement
Figure 2 shows the agreement between reviewers X, Y and Z for all scans and also for each reviewer compared to the reference standard. Good agreement in all cases, \( \kappa > 0.7 \), was found when taking all data into account and this improved to very good agreement in all cases, \( \kappa \geq 0.85 \), for the type A only studies. Agreement was inconsistent for non-type A studies, \(-0.1 < \kappa < 0.9\), with large overlaps in the 95% CI between reviewers. The small sample size of the non-type A studies should be taken note of.

True positive rate (TPR, sensitivity), false positive rate (FPR, 1-specificity) and accuracy for each reader are given in Table 2. All three readers correctly reported type A scans to a high level of accuracy (range = 94–97%) whereas non-type A scans were
reported inconsistently (range = 48–95%). This is supported by the calculated $k$ (±95% CI) scores in Figure 2.

When considering all data the mean accuracy for all 3 readers was 90%, this is equivalent to the average reader accuracy relative to autopsy data from the $^{18F}$-florbetapir trials. 

Individually, readers X and Y are below the 90% average however when the results are sorted by scan type they increase to 96% and 94% respectively for type A scans and the mean accuracy for all readers increases to 96%. For the non-type A scans the mean accuracy for all readers was 70%.

The poor agreement of non-type A scans is predominantly due to the type C subtype (Table 2). All three readers agreed with each other in 7/15 type B cases and 2/6 type C cases (c.f. 72/79 type A). Results for type B and C were not included in the agreement plots due to poor statistical significance as a result of small samples (large 95% CI).

Mean cortical SUVR

ROC curves were plotted for all data, type A and non-type A scans (Figure 3) and the optimal point from each curve was chosen as a new mcSUVR threshold to compare with the $gHC$ threshold and the Pontecorvo mcSUVR threshold$^3$ (Table 3).

The area under the curve (AUC) was greatest for Type A only studies (all data AUC = 0.93, type A = 0.95, non-type A AUC = 0.78) with the optimal point at 1.18. The non-type A curve shows increased FPR compared to the type A curve and has optimal mcSUVR of 1.48. Applying this threshold to the non-type A data yielded an improved MCC of 0.52 compared to both the $gHC$ and Pontecorvo values. There is no difference in classification performance between the ROC optimised type A threshold and the $gHC$ threshold (Table 3).

ROC curves were not generated for type B and C due to the small sample size, however $Aβ+/Aβ−$ classification was assessed with the $gHC$ and Pontecorvo thresholds (Table 3). Each subtype showed the same results for both thresholds however classification of type B scans outperformed classification of type C where the measured FPR was 100%. The poor classification of non-type A is largely controlled by the type C scans.

The agreement across scan sub types (Figure 4) is equivalent to the reader data with best agreement for type A, good agreement for all data and poor to moderate agreement for non-type A. Using the Pontecorvo mcSUVR gives good agreement for all scans and type A in our sample ($\kappa = 0.63±0.15$, $0.75±0.15$ respectively) whereas the $gHC$ and ROC optimised thresholds both show good agreement for all data and very good agreement for type A scans ($gHC \kappa = 0.70±0.14$, $0.85±0.12$; ROC $\kappa = 0.72±0.14$, $0.85±0.12$). For non-type A scans the Pontecorvo threshold is an improvement on the $gHC$ threshold but does not perform as well as the ROC optimised threshold ($\kappa = 0.25±0.37$, $0.22±0.44$ and $0.43±0.44$ respectively, Figure 4).

Two-region classification

ROC curves were plotted for each region of interest in the type A and non-type A scans (Figure 5) and the optimal point from each curve was chosen as a new ROI SUVR threshold to compare with the rHC threshold (Tables 4 and 5).

For non-type A (Figure 5, right) the AUCs range from 0.65 to 0.87, SE = 11%, whereas for the type A studies (Figure 5, left) AUC range is 0.92–0.95, SE = 3%.

Table 2. TPR (sensitivity), FPR (1-specificity) and accuracy for readers against the reference standard. There were no positive type C studies therefore TPR is not applicable.

<table>
<thead>
<tr>
<th></th>
<th>X vs report</th>
<th>Y vs report</th>
<th>Z vs report</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Acc</td>
<td>FPR</td>
<td>TPR</td>
</tr>
<tr>
<td>All</td>
<td>86%</td>
<td>15%</td>
<td>98%</td>
</tr>
<tr>
<td>Type A</td>
<td>96%</td>
<td>2%</td>
<td>97%</td>
</tr>
<tr>
<td>Non-A</td>
<td>54%</td>
<td>48%</td>
<td>100%</td>
</tr>
<tr>
<td>Type B</td>
<td>53%</td>
<td>43%</td>
<td>100%</td>
</tr>
<tr>
<td>Type C</td>
<td>33%</td>
<td>67%</td>
<td>N/A</td>
</tr>
</tbody>
</table>

Acc, accuracy; FPR, false positive rate; TPR, true positive rate.

Figure 3. ROC curves for mcSUVR evaluation of all data, type A and non-type A subgroups.
ROC optimisation improves the accuracy of the two-region classification method in type A scans and non-type A scans (Table 4). Accuracy for the ROC optimised type A studies exceeds the 90% average from the $^{18}$F-florbetapir trials but does not match the accuracy of readers in this study (Table 2). The MCC, TPR and FPR for the type A ROC optimised SUVR thresholds (Table 4) are equivalent to those measured for the gHC and ROC optimised thresholds (Table 3).

Using two-region classification, agreement scores were found to be good for all data and type A and poor for non-type A (Figure 6). This follows the same trend as the reader agreement and mcSUVR agreement. The ROC optimised ROI SUVR improves agreement scores for all data and type A scans compared with the BRASS default values ($\kappa = 0.72\pm0.13$ and $0.85\pm0.12$ respectively) and improves the agreement in non-type A scans from poor to fair ($\kappa = 0.32\pm0.37$).

Similarly to the mcSUVR values, when the type B and C scans are analysed separately it can be seen that the poor performance of non-type A is governed by the type C scans. Classification of type B scans with the rHC thresholds is poorer than with the gHC thresholds (Tables 3 and 4). For both the gHC and rHC thresholds type C scans were all incorrectly classified as false positives (FPR = 100%).

DISCUSSION

The aim of this investigation was to assess and optimise PET imaging derived quantitative thresholds to assist with determining amyloid positivity status in subgroups (according to uptake patterns) of $^{18}$F-florbetapir PET/CT examinations. We have derived new amyloid positivity thresholds, using the methodology available in a commercial software tool approved for clinical use, and compared them against values derived from a healthy control population and a value from another publication where the same software was used.

Type A

For type A scans, readers reported all studies to a high degree of accuracy, equivalent to that reported by other studies. However, here we have observed that division of scans into subtypes can increase reader accuracy and improve inter-reader agreement beyond the levels achieved when considering all scans as a single group. Mean inter-reader accuracy for type A scans was 96% and agreement $\geq 0.85$ in all cases. Furthermore increased sensitivity, specificity and MCC have been observed for both mcSUVR and the two-region classification method for type A scans relative to all scans in a single group.

ROC analysis indicates that using a threshold mcSUVR of 1.18 can achieve 97% TPR and 12% FPR, this ROC derived mcSUVR value is approximately equal to the gHC threshold (1.17). Furthermore equivalent TPR and FPR can be achieved by using the two-region method with optimised ROI SUVR thresholds (Table 4). However as type A studies are defined as having a widespread loss in GMWM differentiation, mcSUVR would be the most suitable metric.

Table 3. Comparison of the gHC threshold, Pontecorvo threshold and ROC optimised threshold.

| n | mcSUVR | TPR | FPR | Acc | MCC | mcSUVR | TPR | FPR | Acc | MCC | mcSUVR | TPR | FPR | Acc | MCC |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| All | 100 | 1.17 | 96% | 24% | 85% | 0.72 | 1.11 | 98% | 2% | 90% | 0.62 | 1.11 | 98% | 2% | 90% | 0.62 |
| Type A | 79 | 1.11 | 97% | 3% | 94% | 0.72 | 1.11 | 98% | 2% | 90% | 0.62 | 1.11 | 98% | 2% | 90% | 0.62 |
| Non-A | 21 | 1.11 | 97% | 3% | 94% | 0.72 | 1.11 | 98% | 2% | 90% | 0.62 | 1.11 | 98% | 2% | 90% | 0.62 |
| Type B | 15 | 1.11 | 97% | 3% | 94% | 0.72 | 1.11 | 98% | 2% | 90% | 0.62 | 1.11 | 98% | 2% | 90% | 0.62 |
| Type C | 6 | N/A | 100% | 0% | -1 | N/A | 100% | 0% | -1 | N/A | 100% | 0% | -1 | N/A | 100% | 0% | -1 |
The ROC optimised mcSUVR for type A scans is equivalent to the mcSUVR from Clark et al\(^8\) which was found to correlate with high probability of AD as confirmed by autopsy. The study obtained this result with a sample of 19 patients; here we have calculated the same threshold with 79 clinical patient scans. This value is also equivalent to the gHC threshold. It is important to note that the healthy controls and the region masks used to calculate the gHC threshold were provided by Avid, the radiopharmaceutical manufacturer, so the database may contain patients used in other published studies which would explain the similarity in threshold values.

For our patient data, the Pontecorvo threshold is outperformed by the gHC threshold and the ROC optimised threshold derived from our sample (Table 3). The decrease in MCC for the Pontecorvo threshold is due to the increase in the FPR for the type A studies suggesting that a threshold of 1.11 is too low. This result corresponds with Schreiber\(^4\) who observed an increase in specificity at threshold 1.17.

One of the limitations of this study (similar to Schreiber and Camus\(^4,6\)) was that the reference standard was the original clinical report with no follow up data to confirm the diagnosis, as opposed to autopsy plaque counts or CSF-\(A\beta\) used in other studies. This may imply that our readers and mcSUVR threshold give a conservative assessment of uptake, however without follow up data from these patients it is not possible to say.

Second, the readers did not use quantitative analysis to augment visual reads as some other studies have done therefore the added benefit can only be hypothesised.

The evidence here suggests that setting the \(A\beta^+\) threshold for type A scans as the gHC threshold (1.17) is a suitable metric
for classification of type A studies. However the sensitivity and specificity is no better than that observed between our readers (Table 2). The supporting information gained from quantification may not provide extra useful information for type A scans but it would not be expected to reduce agreement and accuracy if used as a supporting tool.

Non-type A
Reader accuracy fell to 70% in the non-type A studies (range 48–95%) and agreement ranged from very poor to poor (Figure 2, Table 2). This suggests that readers require further support to assist with reporting these atypical scan sub types. Mean cortical measures may be more difficult to apply to these scans since we expect to see one or more focal regions of uptake, increased noise, motion artefacts and atrophy amongst other features. The ROC mcSUVR does show some improvement relative to the gHC and Pontecorvo values however the sensitivity is 38% which is less than any reader for non-type A studies (Tables 2 and 3). The Pontecorvo threshold shows poor agreement with the reference standard and the specificity is less than any other threshold measured (Table 3) indicating that 1.11 is too low for non-type A studies. MCC values were also found to be consistently poor (Table 3) further supporting the argument against mcSUVR for non-type A scans, thus ROI SUVR would be a more useful metric. The optimum ROC ROI SUVR increased for 5/6 regions relative to type A scans (Table 5), the exception is the precuneus where SUVR decreases. These results suggest that a higher threshold is required for non-type A studies. The decrease in the precuneus is most likely attributable to the small sample size or possible poor region placement. Further investigation is needed. Despite the small sample size, using the ROC derived values does result in some improvement (accuracy and FPR) which is promising for future investigations.

The use of ROI SUVR thresholds for quantification of amyloid brain scans may not be suitable in some cases as individual

<table>
<thead>
<tr>
<th>n</th>
<th>TPR</th>
<th>FPR</th>
<th>Acc</th>
<th>MCC</th>
<th>ROC ROI SUVR</th>
<th>TPR</th>
<th>FPR</th>
<th>Acc</th>
<th>MCC</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>100</td>
<td>98%</td>
<td>33%</td>
<td>81%</td>
<td>0.67</td>
<td>Table 5</td>
<td>97%</td>
<td>12%</td>
<td>92%</td>
</tr>
<tr>
<td>Type A</td>
<td>79</td>
<td>97%</td>
<td>20%</td>
<td>89%</td>
<td>0.79</td>
<td>Table 5</td>
<td>97%</td>
<td>12%</td>
<td>92%</td>
</tr>
<tr>
<td>Non-A</td>
<td>21</td>
<td>100%</td>
<td>77%</td>
<td>52%</td>
<td>0.32</td>
<td>Table 5</td>
<td>75%</td>
<td>38%</td>
<td>61%</td>
</tr>
<tr>
<td>Type B</td>
<td>15</td>
<td>89%</td>
<td>43%</td>
<td>75%</td>
<td>0.49</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Type C</td>
<td>6</td>
<td>N/A</td>
<td>100%</td>
<td>0%</td>
<td>-1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Acc, accuracy; FPR, false positive rate; MCC, Matthews correlation coefficient; ROC, receiver operating characteristic; ROI, region of interest; SUVR, standardised uptake value ratio; TPR, true positive rate.

Table 4. Comparison of two-region classification metrics using control template ROI SUVR and ROC optimised ROI SUVR.

Table 5. rHC thresholds for each ROI and the optimal point from each ROC curve for type A and non-type A studies. FPR* and TPR* values are the ROC co-ordinates from the type A curves (Figure 5, left) for the corresponding values of rHC thresholds. This allows for comparison with the type A ROC optimal values.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>SUVR +2 SD</td>
<td>1.09</td>
<td>1.25</td>
<td>1.21</td>
<td>1.17</td>
<td>1.25</td>
<td>1.05</td>
</tr>
<tr>
<td>FPR*</td>
<td>15%</td>
<td>17%</td>
<td>20%</td>
<td>22%</td>
<td>20%</td>
<td>10%</td>
</tr>
<tr>
<td>TPR*</td>
<td>97%</td>
<td>97%</td>
<td>95%</td>
<td>97%</td>
<td>95%</td>
<td>76%</td>
</tr>
<tr>
<td>Type A ROC optimum threshold</td>
<td>SUVR</td>
<td>1.13</td>
<td>1.25</td>
<td>1.29</td>
<td>1.25</td>
<td>1.38</td>
</tr>
<tr>
<td>FPR</td>
<td>12%</td>
<td>17%</td>
<td>12%</td>
<td>17%</td>
<td>7%</td>
<td>12%</td>
</tr>
<tr>
<td>TPR</td>
<td>97%</td>
<td>97%</td>
<td>95%</td>
<td>95%</td>
<td>87%</td>
<td>92%</td>
</tr>
<tr>
<td>Non-type A ROC optimum threshold</td>
<td>SUVR</td>
<td>1.52</td>
<td>1.38</td>
<td>1.53</td>
<td>1.03</td>
<td>1.52</td>
</tr>
<tr>
<td>FPR</td>
<td>8%</td>
<td>0%</td>
<td>0%</td>
<td>15%</td>
<td>0%</td>
<td>23%</td>
</tr>
<tr>
<td>TPR</td>
<td>25%</td>
<td>50%</td>
<td>50%</td>
<td>75%</td>
<td>63%</td>
<td>88%</td>
</tr>
</tbody>
</table>

Acc, accuracy; Ant. Cing, anterior cingulate; FPR, false positive rate; Front Med. Orb, frontal medial orbital cortex; Lat. Temp, lateral temporal lobes; MCC, Matthews correlation coefficient; Post. Cing, posterior cingulate; ROC, receiver operating characteristic; ROI, region of interest; SUVR, standardised uptake value ratio; TPR, true positive rate.

Single regions are not used to classify scans, the FPR and TPR were used to choose optimum SUVR for each ROI and also for comparison between subgroups and thresholds.
regions are more susceptible to motion errors, particularly smaller regions and distal regions, as well as the possibility of partial volume errors.

A limitation to the non-type A results is the small sample \((n = 21)\), this results in large errors in the ROC data thus these SUVR thresholds are not suitable for use in a clinical setting. The data should be updated and re-evaluated once a larger sample of non-type A studies have been acquired.

Non-type A scans represent 20% of the cohort in this study which was sampled from clinical data, this is a large proportion of patients so improving quantification is a matter that should be addressed. So far no other optimised regional SUVRs have been published for the scan sub types defined and our results suggest that improvements can be made in quantifying these sub types of scans.

Type B and C
Although samples were small some analysis was performed on these subtypes separately, this highlighted that the type C scans were responsible for the poor performance of the non-type A group. The high FPR suggests that the gHC and rHC thresholds were too low but with an improved sample size the classification of type B scans may be improved and allow for optimisation of thresholds. Interestingly, the mcSUVR performed better for classification of type B subjects than the two region method (see definition of type B), however this change in performance only accounts for a change in classification of 1 patient.

With regards to visual reads the agreement with the final report was poor (Table 2) and inter-reader agreement was also poor. This suggests that type B scans may benefit the most from quantification however type C appear the most difficult to report but there is no clear quantitative threshold which will support a clinical decision.

**CONCLUSION**
Reader agreement and accuracy were both observed to be poorer in non-type A compared to type A \(^{18}\text{F}-\text{florbetapir}\) scans. This separation of scans has not been investigated in other studies. Quantification of type A scans (using a threshold mcSUVR of 1.17 or two-region classification with optimised ROI SUVR thresholds) improves sensitivity, specificity and accuracy compared to all data as a single group, however this does not exceed the accuracy of experienced readers for type A scans. Augmenting visual reads with quantitative information can reasonably be expected not to detract from a trained reader’s judgement and could prove to be useful to newly trained, or currently training, readers.

There is scope for significant improvement in the interpretation and quantification of non-type A scans. Quantitative information (with optimised thresholds) may help to improve readers’ confidence and accuracy when reporting non-type A studies, in particular type B.

\(^{18}\text{F}-\text{florbetapir}\) scans should be classified according to sub type before quantification and mcSUVR and ROI SUVR thresholds should be implemented for scan sub types.

**ACKNOWLEDGEMENTS**
Dr P Malhotra is supported by the NIHR Imperial Biomedical Research Centre.
REFERENCES


